• What is time series data

• Estimating a causal effect vs forecasting

• Lags, first differences and growth rates

• Autocorrelation

• Autoregressions

• Auto regressive distributed lag model

• Nonstationarity: stochastic trends
  • random walk with and without drift
  • testing for stochastic trends (Dickey-Fuller test)
What is time series data?

Cross section data is data collected for multiple entities at one point in time.

Panel data is data collected for multiple entities at multiple points in time.

Time series data is data collected for a single entity at multiple points in time:

- Yearly GDP of Norway for a period of 20 years
- Daily NOK/Euro exchange rate for past year
- Quarterly data on the inflation & unemployment rate in the U.S from 1957-2005
What is time series data?

Quarterly time series data on inflation for the U.S. from 1957 to 2005
What is time series data?

Quarterly time series data on unemployment for the U.S. from 1957-2005
What is time series data?

Quarterly data on inflation & unemployment in the U.S. 2000-2005:

<table>
<thead>
<tr>
<th>time</th>
<th>inflation</th>
<th>unemployme-e</th>
</tr>
</thead>
<tbody>
<tr>
<td>173</td>
<td>2000q1</td>
<td>3.9386022</td>
</tr>
<tr>
<td>174</td>
<td>2000q2</td>
<td>3.1231739</td>
</tr>
<tr>
<td>175</td>
<td>2000q3</td>
<td>3.6388498</td>
</tr>
<tr>
<td>176</td>
<td>2000q4</td>
<td>2.8415304</td>
</tr>
<tr>
<td>177</td>
<td>2001q1</td>
<td>3.8089782</td>
</tr>
<tr>
<td>178</td>
<td>2001q2</td>
<td>3.0958201</td>
</tr>
<tr>
<td>179</td>
<td>2001q3</td>
<td>0.9015987</td>
</tr>
<tr>
<td>180</td>
<td>2001q4</td>
<td>-0.5256825</td>
</tr>
<tr>
<td>181</td>
<td>2002q1</td>
<td>1.4252196</td>
</tr>
<tr>
<td>182</td>
<td>2002q2</td>
<td>3.2068807</td>
</tr>
<tr>
<td>183</td>
<td>2002q3</td>
<td>2.0744651</td>
</tr>
<tr>
<td>184</td>
<td>2002q4</td>
<td>2.0637621</td>
</tr>
<tr>
<td>185</td>
<td>2003q1</td>
<td>4.0958529</td>
</tr>
<tr>
<td>186</td>
<td>2003q2</td>
<td>0.36346725</td>
</tr>
<tr>
<td>187</td>
<td>2003q3</td>
<td>2.1751006</td>
</tr>
<tr>
<td>188</td>
<td>2003q4</td>
<td>0.86675908</td>
</tr>
<tr>
<td>189</td>
<td>2004q1</td>
<td>3.8057824</td>
</tr>
<tr>
<td>190</td>
<td>2004q2</td>
<td>4.3359083</td>
</tr>
<tr>
<td>191</td>
<td>2004q3</td>
<td>1.6227116</td>
</tr>
<tr>
<td>192</td>
<td>2004q4</td>
<td>3.5051154</td>
</tr>
<tr>
<td>193</td>
<td>2005q1</td>
<td>2.3660429</td>
</tr>
</tbody>
</table>
What is time series data?

- A particular observation $Y_t$ indexed by subscript $t$
- Total number of observations equals $T$
- $Y_t$ is current value and value in previous period is $Y_{t-1}$ (first lag)
- In general $Y_{t-j}$ is called $j$th lag and similarly, $Y_{t+j}$ is the $j$th future value
- The first difference $Y_t - Y_{t-1}$ is the change in $Y$ from period $t - 1$ to period $t$
- Time series regression models can be used for
  - (1) estimating (dynamic) causal effects;
  - (2) forecasting
Estimating (dynamic) causal effect vs forecasting

- Time series data is often used for forecasting
  - For example next year’s economic growth is forecasted based on past and current values of growth & other (lagged) explanatory variables
- Forecasting is quite different from estimating causal effects and is generally based on different assumptions.
- Models that are useful for forecasting need not have a causal interpretation!
  - OLS coefficients need not be unbiased & consistent
- Measures of fit, such as the (adjusted) $R^2$ or the SER
  - are not very informative when estimating causal effects
  - are informative about the quality of a forecasting model
Logarithms and growth rates

- Suppose that $Y_t$ is some time series, then the rate of growth of $Y$ is
  \[ \frac{Y_t - Y_{t-1}}{Y_{t-1}} = \frac{\Delta Y_t}{Y_{t-1}} \]

- instead, we often use the “logarithmic growth" or “log-difference":
  \[
  \Delta \ln Y_t = \ln(Y_t) - \ln(Y_{t-1})
  = \ln \left( \frac{Y_t}{Y_{t-1}} \right)
  = \ln \left( \frac{Y_t}{Y_{t-1}} + \frac{Y_{t-1}}{Y_{t-1}} - \frac{Y_{t-1}}{Y_{t-1}} \right)
  = \ln \left(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}}\right) \approx \frac{\Delta Y_t}{Y_{t-1}}
  \]

see also formula (8.16) of S&W
Annualized rate of inflation:

\[ inflation_t \approx 100 \times 4 \times (\ln(Y_t) - \ln(Y_{t-1})) \]

Stata:

```stata
tsset time
gen ln_CPI=ln(CPI)
gen ln_CPI_1stlag=ln(L1.CPI)
gen inflation=400*(ln(CPI) - ln(L1.CPI))
```

<table>
<thead>
<tr>
<th>time</th>
<th>CPI</th>
<th>ln(CPI)</th>
<th>ln(CPI)_1stlag</th>
<th>inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1957q1</td>
<td>27.77667</td>
<td>3.3241963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1957q2</td>
<td>28.01333</td>
<td>3.3326806</td>
<td>3.3241963</td>
<td>3.3937128</td>
</tr>
<tr>
<td>1957q3</td>
<td>28.26333</td>
<td>3.3415654</td>
<td>3.3326806</td>
<td>3.553894</td>
</tr>
<tr>
<td>1957q4</td>
<td>28.4</td>
<td>3.3463891</td>
<td>3.3415654</td>
<td>1.9295103</td>
</tr>
<tr>
<td>1958q1</td>
<td>28.73667</td>
<td>3.3581739</td>
<td>3.3463891</td>
<td>4.7138908</td>
</tr>
<tr>
<td>1958q2</td>
<td>28.93</td>
<td>3.3648791</td>
<td>3.3581739</td>
<td>2.6821083</td>
</tr>
<tr>
<td>1958q3</td>
<td>28.91333</td>
<td>3.3643029</td>
<td>3.3648791</td>
<td>-0.23050418</td>
</tr>
<tr>
<td>1958q4</td>
<td>28.94333</td>
<td>3.3653399</td>
<td>3.3643029</td>
<td>0.41480139</td>
</tr>
</tbody>
</table>
In time series data, $Y_t$ is typically correlated with $Y_{t-j}$, this is called \textbf{autocorrelation} or \textbf{serial correlation}.

- The $j^{\text{th}}$ autocovariance $= \text{Cov}(Y_t, Y_{t-j})$ can be estimated by
  \[
  \text{Cov}(\overline{Y_t}, \overline{Y_{t-j}}) = \frac{1}{T} \sum_{t=j+1}^{T} (Y_t - \overline{Y_{j+1,T}})(Y_{t-j} - \overline{Y_{1,T-j}})
  \]
  where $\overline{Y_{j+1,T}}$ is the sample average of $Y$ computed over observations $t = j+1, \ldots, T$
  and $\overline{Y_{1,T-j}}$ is the sample average of $Y$ computed over observations $t = 1, \ldots, T-j$.

- The $j^{\text{th}}$ autocorrelation $= \rho_j = \frac{\text{Cov}(Y_t, Y_{t-j})}{\text{Var}(Y_t)}$ can be estimated by
  \[
  \hat{\rho}_j = \frac{\text{Cov}(\overline{Y_t}, \overline{Y_{t-j}})}{\text{Var}(Y_t)}
  \]
  $j$ start-up observations are lost in constructing these sample statistics.

- Denominator of $j^{\text{th}}$ autocorrelation assumes stationarity of $Y_t$, which implies (among other things) $\text{Var}(Y_t) = \text{Var}(Y_{t-j})$.
First 4 autocorrelations of inflation ($\text{inf}_t$):

```
. corrgram inflation if tin(1960q1,2004q4), noplot lags(4)

<table>
<thead>
<tr>
<th>LAG</th>
<th>AC</th>
<th>PAC</th>
<th>Q</th>
<th>Prob&gt;Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8359</td>
<td>0.8361</td>
<td>127.89</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.7575</td>
<td>0.1937</td>
<td>233.49</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.7598</td>
<td>0.3206</td>
<td>340.34</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>0.6699</td>
<td>-0.1881</td>
<td>423.87</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
```

First 4 autocorrelations of the change in inflation ($\triangle \text{inf}_t = \text{inf}_t - \text{inf}_{t-1}$):

```
. corrgram D.inflation if tin(1960q1,2004q4), noplot lags(4)

<table>
<thead>
<tr>
<th>LAG</th>
<th>AC</th>
<th>PAC</th>
<th>Q</th>
<th>Prob&gt;Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.2618</td>
<td>-0.2636</td>
<td>12.548</td>
<td>0.0004</td>
</tr>
<tr>
<td>2</td>
<td>-0.2549</td>
<td>-0.3497</td>
<td>24.507</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.2938</td>
<td>0.1461</td>
<td>40.481</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>-0.0605</td>
<td>-0.0220</td>
<td>41.162</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
```
Stationarity

• The denominator of $j^{th}$ autocorrelation assumes stationarity of $Y_t$

• A time series $Y_t$ is stationary if its probability distribution does not change over time,
  - when the joint distribution of $(Y_{s+1}, \ldots, Y_{s+T})$ does not depend on $s$

• Stationarity implies that $Y_1$ has the same distribution as $Y_t$ for any $t = 1, 2, \ldots$

• In other words, $\{Y_1, Y_2, \ldots, Y_T\}$ are identically distributed, however, they are not necessarily independent!

• If a series is nonstationary, then conventional hypothesis tests, confidence intervals and forecasts can be unreliable.

• Stationarity says that history is relevant, it is a key requirement for external validity of time series regression.
First order autoregressive model: AR(1)

- Suppose we want to forecast the change in inflation from this quarter to the next.
- When predicting the future of a time series a good place to start is in the immediate past.
- The first order autoregressive model (AR(1))

\[ Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t \]

- Forecast in next period based on AR(1) model:

\[ \hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T \]

- Forecast error is the mistake made by the forecast

\[ Forecast\ error = Y_{T+1} - \hat{Y}_{T+1|T} \]
Forecast vs predicted value & forecast error vs residual

- A forecast is not the same as a predicted value
- A forecast error is not the same as a residual

OLS predicted values $\hat{Y}_t$ and residuals $\hat{u}_t = Y_t - \hat{Y}_t$ for $t \leq T$ are “in-sample”:

- They are calculated for the observations in the sample used to estimate the regression.
  - $Y_t$ is observed in the data set used to estimate the regression.

A forecast $\hat{Y}_{T+j|T}$ and forecast error $Y_{T+j} - \hat{Y}_{T+j|T}$ for $j \geq 1$ are “out-of-sample”:

- They are calculated for some date beyond the data set used to estimate the regression.
  - $Y_{T+j}$ is not observed in the data set used to estimate the regression.
First order autoregressive model: AR(1)

\[ \Delta \text{inflation}_t = \beta_0 + \beta_1 \Delta \text{inflation}_{t-1} + u_t \]

> gen d_inflation=D1.inflation
(2 missing values generated)

1. regress d_inflation L1.d_inflation if tin(1962q1,2004q4), r

Linear regression                                      Number of obs = 172
F(  1,   170) = 6.08
Prob > F      = 0.0146
R-squared     = 0.0564
Root MSE      = 1.664

|                    | Coef.  | Std. Err. | t     |  P>|t| | [95% Conf. Interval] |
|--------------------|--------|-----------|-------|------|---------------------|
| d_inflation        |        |           |       |      |                     |
| L1.                | -.2380471 | .0965017   | -2.47 | 0.015 | -.4285431 - .047551 |
| _cons              | .0171008 | .1268849   | 0.13  | 0.893 | -.2333721 .2675736 |

2. dis "Adjusted Rsquared = " _result(8)
   Adjusted Rsquared = .05082857

- \( \triangle \text{inf}_{2005q1|2004q4} = 0.017 - 0.238 \cdot \triangle \text{inf}_{2004q4} = -0.43 \)

- Forecast error = \( \triangle \text{inf}_{2005q1} - \hat{\triangle \text{inf}}_{2005q1|2004q4} = -1.14 - (-0.43) = -0.71 \)
Forecasts are uncertain and the Root Mean Squared Forecast Error (RMSFE) is a measure of forecast uncertainty.

The RMSFE is a measure of the spread of the forecast error distribution.

\[
RMSFE = \sqrt{E \left[ \left( Y_{T+1} - \hat{Y}_{T+1|T} \right)^2 \right]}
\]

The RMSFE has two sources of error:

1. The error arising because future values of \( u_t \) are unknown
2. The error in estimating the coefficients \( \beta_0 \) and \( \beta_1 \)

If the sample size is large the first source of error will be larger than the second and \( RMSFE \approx \sqrt{Var(\hat{u}_t)} \)

\( \sqrt{Var(\hat{u}_t)} \) can be estimated by the \( SER = \frac{1}{T-2} \sum_{t=1}^{T} \hat{u}_t^2 \).
The \textit{pth} order autoregressive model (AR(p)) is

\[ Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t \]

- The AR(p) model uses \( p \) lags of \( Y \) as regressors.
- The number of lags \( p \) is called the order or lag length of the autoregression.
- The coefficients generally do not have a causal interpretation.
- We can use t- or F-tests to determine the lag order \( p \).
- Or we can determine \( p \) using an “information criterion” (more on this later...)
\[ \text{AR(4)} \]

> `regress d_inflation L1.d_inflation L2.d_inflation L3.d_inflation L4.d_inflation if tin(1962q1,2004q4), r`

Linear regression

\[
\begin{array}{cccccc}
\text{d_inflation} & \text{Coef.} & \text{Std. Err.} & t & P>|t| & [95\% \text{ Conf. Interval}] \\
\text{d_inflation} & & & & & \\
\text{L1.} & -0.2579426 & 0.0925934 & -2.79 & 0.006 & -0.4407471 -0.0751381 \\
\text{L2.} & -0.3220312 & 0.0805465 & -4.00 & 0.000 & -0.4810518 -0.1630106 \\
\text{L3.} & 0.1576089 & 0.0841017 & 1.87 & 0.063 & -0.0084307 0.3236484 \\
\text{L4.} & -0.0302511 & 0.0930471 & -0.33 & 0.746 & -0.2139512 0.153449 \\
\text{_cons} & 0.0224294 & 0.1176344 & 0.19 & 0.849 & -0.2098127 0.2546715 \\
\end{array}
\]

\[ \text{Adjusted Rsquared = } 0.18475367 \]

- Adjusted \( R^2 \) is higher and the RMSE is lower in AR(4) than in AR(1)

\[ \hat{\Delta inf}_{2005q1|2004q4} \approx 0.41 \]

- Forecast error = \( \Delta inf_{2005q1} - \hat{\Delta inf}_{2005q1|2004q4} = -1.14 - (0.4) = -1.55 \]
Lag length selection

• Is the AR(4) model better that the AR(1) model?

. test L2.d_inflation=L3.d_inflation=L4.d_inflation=0

( 1)  L2.d_inflation - L3.d_inflation = 0
( 2)  L2.d_inflation - L4.d_inflation = 0
( 3)  L2.d_inflation = 0

\[ F(3, 167) = 6.71 \]
\[ \text{Prob} > F = 0.0003 \]

How should we choose the lag length \( p \)?

• One approach is to start with a model with many lags and to perform a hypothesis test on the final lag

• Delete the final lag if insignificant and perform an hypothesis test on the new final lag, ..., continue until all included lags are significant.

• Drawback of this approach is that it can produce too large a model

  • at a 5% significance level: if the true lag length is 5 it will estimate \( p \) to be 6 in 5% of the time.
Lag length selection

Alternative way to determine the lag length $p$ is to minimize one of the following information criteria:

- **Bayes information criterion (BIC):**
  
  $$BIC(p) = \ln \left[ \frac{SSR(p)}{T} \right] + (p + 1) \frac{\ln(T)}{T}$$

- **Akaike information criterion (AIC):**
  
  $$AIC(p) = \ln \left[ \frac{SSR(p)}{T} \right] + (p + 1) \frac{2}{T}$$

- $SSR(p)$ is $\sum_{t=1}^{T} \hat{u}_t^2$ in an AR($p$) model

- $T$ is the number of time periods

- In order to compare the BIC (or AIC) for different $p$,...

- ....all autoregressions with different lag lengths $p$ should be based on the same number of observations $T$!
Lag length selection

The BIC and AIC both consist of two terms:

- 1st term $\ln \left( \frac{SSR(p)}{T} \right)$: always decreasing in $p$
  - larger $p$, better fit

- 2nd term $(p + 1) \frac{\ln(T)}{T}$ (BIC) or $(p + 1) \frac{2}{T}$ (AIC): always increasing in $p$
  - This term is a “penalty” for estimating more parameters – and thus increasing the RMSFE.

AIC estimates more lags (larger $p$) than the BIC for $T > 7.4$,

- the penalty term is smaller for AIC than BIC

In large samples the AIC overestimates $p$, it is inconsistent.
Lag length selection

1. `regress d_inflation L1.d_inflation if tin(1962q1,2004q4), r`

Linear regression  
Number of obs = 172
F( 1, 170) = 6.08
Prob > F = 0.0146
R-squared = 0.0564
Root MSE = 1.664

|                  | Coef. | Std. Err. | t     | P>|t|      | [95% Conf. Interval] |
|------------------|-------|-----------|-------|----------|----------------------|
| d_inflation      |       |           |       |          |                      |
| L1.              | -.2380471 | .0965017  | -2.47 | 0.015    | -.4285431            |
| _cons            | .0171008  | .1268849  | 0.13  | 0.893    | -.2333721            |

2. `gen BIC_1=ln(e(rss)/e(N))+e(rank)*(ln(e(N))/e(N)) if tin(1962q1,2004q4)`  
(21 missing values generated)

3. `gen AIC_1=ln(e(rss)/e(N))+e(rank)*(2/e(N)) if tin(1962q1,2004q4)`  
(21 missing values generated)

4. `sum BIC_1 AIC_1`

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC_1</td>
<td>172</td>
<td>1.066562</td>
<td>0</td>
<td>1.066562</td>
<td>1.066562</td>
</tr>
<tr>
<td>AIC_1</td>
<td>172</td>
<td>1.029963</td>
<td>0</td>
<td>1.029963</td>
<td>1.029963</td>
</tr>
</tbody>
</table>
Lag length selection

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC_0</td>
<td>172</td>
<td>1.094665</td>
</tr>
<tr>
<td>BIC_1</td>
<td>172</td>
<td>1.066562</td>
</tr>
<tr>
<td>BIC_2</td>
<td>172</td>
<td>.9549263</td>
</tr>
<tr>
<td>BIC_3</td>
<td>172</td>
<td>.9574141</td>
</tr>
<tr>
<td>BIC_4</td>
<td>172</td>
<td>.9864399</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC_0</td>
<td>172</td>
<td>1.076366</td>
</tr>
<tr>
<td>AIC_1</td>
<td>172</td>
<td>1.029963</td>
</tr>
<tr>
<td>AIC_2</td>
<td>172</td>
<td>.9000281</td>
</tr>
<tr>
<td>AIC_3</td>
<td>172</td>
<td>.8842165</td>
</tr>
<tr>
<td>AIC_4</td>
<td>172</td>
<td>.894943</td>
</tr>
</tbody>
</table>

- Optimal lag length according to BIC: $p = 2$
- Optimal lag length according to AIC: $p = 3$
• Economic theory often suggests other variables that could help to forecast the variable of interest.

• When we add other variables and their lags the result is an autoregressive distributed lag model ADL(p,q)

\[ Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} + \delta_1 X_{t-1} + ... + \delta_q X_{t-q} + u_t \]

• \( p \) is the number of lags of the dependent variable

• \( q \) is the number of lags (distributed lags) of the additional predictor \( X \)
• When predicting future changes in inflation economic theory suggests that lagged values of the unemployment rate might be a good predictor.

• Short-run Philips curve: negative short run relation between unemployment and inflation.

• According to Bays Information Criteria we should include 2 lags of the dependent variable.

• In addition we include 2 lags of the unemployment rate.

• This gives an ADL(2,2) model.
> regress d_inflation L1.d_inflation L2.d_inflation L1.unemployment L2.unemployment if tin(1962q1,2004q4), r

| d_inflation  | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------------|--------|-----------|-------|-------|----------------------|
| d_inflation  |        |           |       |       |                      |
| L1.          | -.4685035 | .0771115  | -6.08 | 0.000 | -.6207425            |
| L2.          | -.4251441 | .0821394  | -5.18 | 0.000 | -.5873096            |
| unemployment_rate |        |           |       |       |                      |
| L1.          | -2.243865 | .4020402  | -5.58 | 0.000 | -3.037602            |
| L2.          | 2.044221  | .3875693  | 5.27  | 0.000 | 1.279054             |
| _cons        | 1.193032  | .4344711  | 2.75  | 0.007 | .3352683             |

\[ \text{Adjusted Rsquared} = \text{.3358903} \]

\begin{itemize}
  \item Adjusted \( R^2 \) is higher and the RMSE is lower in ADL(2,2) than in AR(4)
  \item Forecast error \( = \triangle \text{inf}_{2005q1} - \hat{\triangle \text{inf}}_{2005q1|2004q4} = -1.14 - (0.38) = -1.52 \)
\end{itemize}
Granger “causality” test

- Do the included lags of unemployment have useful predictive content conditional on the included lags of the change in inflation?

- The claim that a variable has no predictive content corresponds to the null hypothesis that the coefficients on all lags of the variable are zero.

- The F-statistic of this test is called the Granger causality statistic.

- If the null hypothesis is rejected the variable $X$ is said to Granger-cause the dependent variable $Y$.

- This does not mean that we have estimated the causal effect of $X$ on $Y$!!

- It means that $X$ is a useful predictor of $Y$ (Granger predictability would be a better term).
The Granger causality test in the ADL(2,2) model:

```
. regress d_inflation L1.d_inflation L2.d_inflation L1.unemployment L2.unemployment if > tin(1962q1,2004q4), r noheader
```

|                  | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|------------------|--------|-----------|-------|------|---------------------|
| **d_inflation**  |        |           |       |      |                     |
| L1.              | -.4685035 | .0771115 | -6.08 | 0.000 | -.6207425 to -.3162645 |
| L2.              | -.4251441 | .0821394 | -5.18 | 0.000 | -.5873096 to -.2629787 |
| **unemployment_rate**  |        |           |       |      |                     |
| L1.              | -2.243865 | .4020402 | -5.58 | 0.000 | -3.037602 to -1.450129 |
| L2.              | 2.044221  | .3875693 | 5.27  | 0.000 | 1.279054 to 2.809388  |
| _cons            | 1.193032  | .4344711 | 2.75  | 0.007 | .3352683 to 2.050796  |

```
. test L1.unemployment=L2.unemployment=0
```

( 1)  L.unemployment_rate - L2.unemployment_rate = 0
( 2)  L.unemployment_rate = 0

\[
F( 2, 167) = 16.13 \\
\text{Prob} > F = 0.0000
\]

- Null hypothesis that coefficients on the 2 lags of unemployment are zero is rejected at a 1% level.

- Unemployment is a useful predictor for the change in the inflation rate.
Nonstationarity: trends

- A time series $Y_t$ is stationary if its probability distribution does not change over time.
- If a time series has a trend, it is nonstationary.
- A trend is a persistent long-term movement of a variable over time.
- We consider two types of trends:
  
  **Deterministic trend:** $Y_t = \beta_0 + \lambda t + u_1$
  - series is a nonrandom function of time
  
  **Stochastic trend:** $Y_t = \beta_0 + Y_{t-1} + u_1$
  - series is a random function of time
Nonstationarity: trends

Deterministic trend

Stochastic trend
Simplest model of a variable with a stochastic trend is the **random walk**

\[ Y_t = Y_{t-1} + u_t \quad \text{where } u_t \text{ is i.i.d.} \]

The value of the series tomorrow is its value today plus an unpredictable change.

An extension of the random walk model is the **random walk with drift**

\[ Y_t = \beta_0 + Y_{t-1} + u_t \quad \text{where } u_t \text{ is i.i.d.} \]

\( \beta_0 \) is the “drift” of the random walk, if \( \beta_0 \) is positive \( Y_t \) increases on average.

A random walk is nonstationary: the distribution is not constant over time.

The variance of a random walk increases over time:

\[ \text{Var} (Y_t) = \text{Var} (Y_{t-1}) + \text{Var} (u_t) \]
• The random walk model is a special case of an AR(1) model with \( \beta_1 = 1 \)

• If \( Y_t \) follows and AR(1) with \( |\beta_1| < 1 \) (and \( u_t \) is stationary), \( Y_t \) is stationary

• If \( Y_t \) follows and AR(p) model

\[
Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t
\]

\( Y_t \) is stationary if its roots \( z \) are all greater than 1 in absolute value.

• the roots are the values of \( z \) that satisfy

\[
1 - \beta_1 z - \beta_2 z^2 - \ldots - \beta_p z^p = 0
\]

• In the special case of an AR(1): \( 1 - \beta_1 z = 0 \) gives \( z = \frac{1}{\beta_1} \) which is bigger than \( 1 \) if \( |\beta_1| < 1 \)
Detecting stochastic trends: Dickey-Fuller test in AR(1) model

- Trends can be detected by informal and formal methods
  - **Informal**: inspect the time series plot
  - **Formal**: Perform the Dickey-Fuller test to test for a stochastic trend.

- Dickey-Fuller test in an AR(1) model:
  
  \[H_0 : \beta_1 = 1 \quad \text{vs} \quad H_1 : \beta_1 < 1 \quad \text{in} \quad Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t\]

- Test can be performed by adding and subtracting \(Y_{t-1}\) from both sides of the equation and estimate
  
  \[\Delta Y_t = \beta_0 + \delta Y_{t-1} + \varepsilon_t\]

- We can now test
  
  \[H_0 : \delta = \beta_1 - 1 = 0 \quad \text{vs} \quad H_1 : \delta < 0\]
Does U.S. inflation have a stochastic trend?

- DF test for a unit root in U.S. inflation (*Note:* we test for a stochastic trend in \( \text{inflation}_t \) and not in \( \Delta \text{inflation}_t \))

```
1. regress d_inflation L1.inflation if tin(1962q1,2004q4), noheader
```

| d_inflation | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-------------|-------|-----------|------|-----|---------------------|
| inflation   | -.1643553 | .0415311 | -3.96 | 0.000 | -.2463383 -.0823722 |
| L1.         |        |           |      |      |                     |
| _cons       | .7219345 | .217599   | 3.32 | 0.001 | .2923904 1.151479   |

- Under \( H_0 \), \( Y_t \) is nonstationary and the DF-statistic has a nonnormal distribution, we therefore use the following critical values

<table>
<thead>
<tr>
<th>TABLE 14.5</th>
<th>Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic Regressors</td>
<td>10%</td>
</tr>
<tr>
<td>Intercept only</td>
<td>-2.57</td>
</tr>
<tr>
<td>Intercept and time trend</td>
<td>-3.12</td>
</tr>
</tbody>
</table>

- \( DF = -3.96 \) this is more negative than -2.86 so we reject the null hypothesis of a stochastic trend at a 5% significance level.
• Often an AR(1) model does not capture all the serial correlation in $Y_t$ and we should include more lags & estimate an AR(p) model.

• We can test for a stochastic trend in an AR(p) model by augmenting the DF-regression by lags of $\Delta Y_t$.

• The Augmented Dickey-Fuller test:

$$H_0 : \delta = 0 \quad vs \quad H_1 : \delta < 0$$

in the regression

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_p \Delta Y_{t-p} + u_t$$

• **Note:** The DF-statistic should be computed using homoskedasticity-only (nonrobust) standard errors (see footnote 3 in S&W CH14)
Does U.S. inflation have a stochastic trend?

- DF test for a unit root in U.S. inflation – using \( p = 4 \) lags (AR(4)-model)

```
1. regress d_inflation L1.inflation L1.d_inflation L2.d_inflation L3.d_inflation L4.d_inflation if tin(1962q1,2004q4), noheader
```

| d_inflation | Coef.   | Std. Err. | t       | P>|t|     | [95% Conf. Interval] |
|--------------|---------|-----------|---------|---------|---------------------|
| inflation    |         |           |         |         |                     |
| L1.          | -.1134169 | .0422344  | -2.69   | 0.008   | -.1968029 -.030031  |
| d_inflation  |         |           |         |         |                     |
| L1.          | -.1864426 | .0805144  | -2.32   | 0.022   | -.3454068 -.0274783 |
| L2.          | -.2563879 | .081463   | -3.15   | 0.002   | -.4172251 -.0955507 |
| L3.          | .1990491  | .0793514  | 2.51    | 0.013   | .0423811 .3557171   |
| L4.          | .0099994  | .0779921  | 0.13    | 0.898   | -.1439849 .1639837  |
| _cons        | .5068158  | .2141807  | 2.37    | 0.019   | .0839466 .9296851   |

- \( DF = -2.69 \) this is less negative than -2.86 so we do not reject the null hypothesis of a stochastic trend at a 5% significance level.
Does U.S. inflation have a stochastic trend?

- Instead of testing the null hypothesis of a stochastic trend against the alternative hypothesis of no trend....

- ...the alternative hypothesis can be that $Y_t$ is stationary around a deterministic trend.

- The Dickey-Fuller regression then includes a deterministic trend

  $$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_p \Delta Y_{t-p} + u_t$$

- And we have to use the critical values in the second row:

<table>
<thead>
<tr>
<th>Deterministic Regressors</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept only</td>
<td>−2.57</td>
<td>−2.86</td>
<td>−3.43</td>
</tr>
<tr>
<td>Intercept and time trend</td>
<td>−3.12</td>
<td>−3.41</td>
<td>−3.96</td>
</tr>
</tbody>
</table>
Avoiding problems caused by stochastic trends

- Best way to deal with a trend is to transform the series such that it does not have a trend.
- If the series $Y_t$ has a stochastic trend, then the first difference of the series $\Delta Y_t$ does not have a stochastic trend.
- For example if $Y_t$ follows a random walk with drift

$$Y_t = \beta_0 + Y_{t-1} + u_t$$

the first difference is stationary

$$\Delta Y_t = \beta_0 + u_t$$

- On slide 37 we saw that we did not reject the null hypothesis of a stochastic trend in $inflation_t$.
- This is the reason that in the beginning of the lecture we estimated AR’s with $\Delta inflation_t$. 